Math 203

Topics for the Chapter 10 quiz:
Probability

To do:

Describe the probability model for a trial: sample space, oucomes, probabilities.

Compute probablilities of events for a trial with equally likely outcomes.

Construct a weighted rooted tree diagram for a multistage trial, compute probabilities of events.

Compute the conditional probability of an event.

Compute the expected value of a probability model.

Probability is the study of the long-term behavior of random phenomena, where *random*, basically, means that knowledge of what the phenomenon has done before will not let you decide what it will do ňext. Such phenomena include flipping coins, rolling dice, the Dow Jones Industrial Average, etc. The basic idea is that while the object's short term behavior is impossible to predict, its long term (average) behavior can be predicted with great accuracy!
Each observation of the object is a *trial* (e.g., the flip of a coin; and each possible outcome of the trial is an *event*. The *probability* of each event predicts how many times the event will occur, in a large number of trials.
We can express these things in a *probability model*. It consists of two things:

1. A *sample space* S = the collection of all possible outcomes for our trial
2. A *probability* (= a number between 0 and 1) for each outcome.

The idea is that the probability describes the fraction of times we would ĕxpect our outcome to occur in a very large number of trials.
The individual probablilities must add up to one, because: If we let an *event* mean, more generally, some collection of outcomes, then the probability of the event should be the sum of the individual probabilities of each event. Consequently, the sum od all the probabilities should be the probability that some one of the outcomes occurs in each trial, i.e., the fraction of the time that something happens! Since šomething always happens, this probability is one.

Ex: flipping a (fair!) coin; the sample space is {heads,tails}, and each has a probability of .5 .
Ex: rolling a pair of dice: there are 36 possible outcomes (if we keep track of which die is which), each having a probability of 1/36 .
These probability models describe *equally likely outcomes*; each outcome has the same probability, which add up to 1; so each outcome has probability 1/(the size of the sample space).

Basic properties of probability models:
A, B = events = collections of possible outcomes; $\overline{A}$ = complement of A = all possible outcomes not in A. $P(A)$ = probability of event A. $A \cup B$ (= union) = the outcomes that are in A <u>or</u> B (or both). $A \cap B$ (= intersection) = the outcomes that are in both A <u>and</u> B. Then:

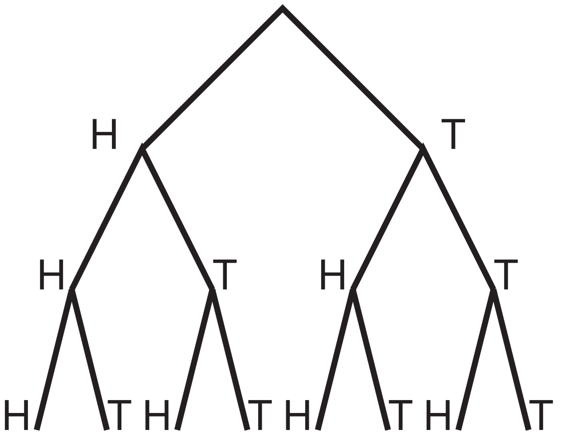
$0 \leq P(A) \leq 1$
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$P(\overline{A}) = 1 - P(A)$
$P(S) = 1$ (in a trial, <u>something</u> happens...)

Two events are *mutually exclusive* if $A \cap B$ is empty, i.e., they share no outcome in common. For mutually exclusive events, we have $P(A \cup B) = P(A) + P(B)$. Different outcomes are mutually exclusive, so the probability of an event = the sum of the probabilities of each individual outcome in the event.

Tree diagrams:

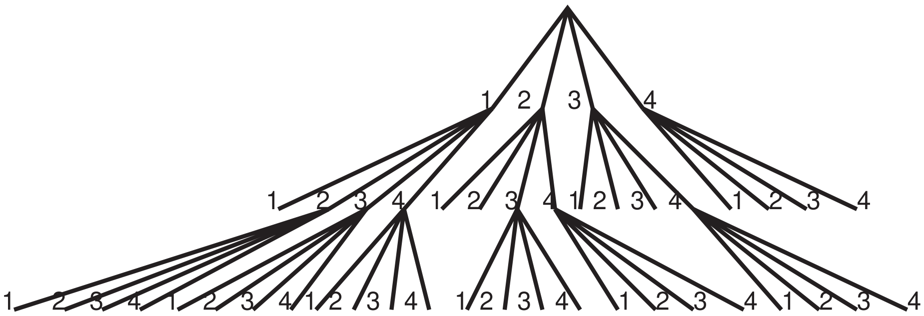
Trees can help us keep track of the relationships among outcomes, and compute the probabilities of events, especially for *multistage trials*: A trial which is a combination of simpler (to compute the probability of!) trials occuring one after the other. This includes the possiblity that what is done in a later stage depends on the outcome of an earlier stage. Examples: (a) draw three balls from a jar (without replacing them between each stage); (b) flip a coin, if heads, roll a die, if tails spin a wheel; (c) roll a 4-sided die and then spin a wheel with numbers 1 through 4, if the die is smaller than the wheel, roll the die again, then take the sum of the die and the wheel as outcome.

The tree describes all of the possible outcomes at each stage of the trial. For example, flipping a coin three times:



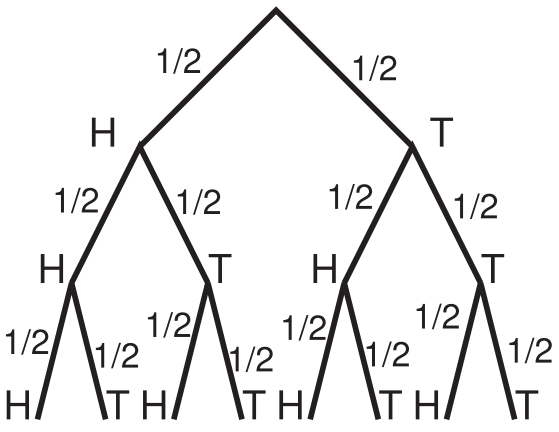
The top vertex is the *root*, the start of the trial. for each successive stage of the trial, we send out *branches* for each of the possible outcomes of the stage, from each of the outcomes of the previous stage that the next stage acts upon. At the bottom are the *leaves*, which represent the posible outcomes of the multi-stage trial. For a 3 coin flip, there are 8 possible outcomes. What each one is can be determined by reading outcomes for each stage from the root to the leaf. For example, the fourth leaf from the left is the outcome HTT.

The tree diagram for example (c) is:



The vertices on the second row with no branches out of them are also leaves.

The tree diagram allows us to see what the possible outcomes of a multi-stage trial are, but more than that, it allows us to compute the probabilities of those outcomes. The idea is that each stage consists of a "simpler" trial, whose probabilities we can compute. We can label each branch with the probability that, at that stage, the outcome at its end will occur. For example, with the 3 coin flip of a (fair) coin:



Each label is 1/2, since each flip is fair. Then the probability of each leaf (that is, of the multi-stage trial ending with the outcome represented by that leaf) is the <u>product</u> of the probabilities on the branches leading to that leaf. This is because if we think in terms of a large number of these multistage trials, if the labels leading to the leaf are $p_1, \ldots, p_k$, then p_1 of the time, the first stage will lead us down the first branch leading to our leaf (that's what the p_1 means!); of those times, p_2 of the time we will go down the second branch, so p_2

<u>of</u> p_1 of the time (i.e., $p_1 p_2$ of the time!), the trial leads us down to the end of the second branch. Repeating this reasoning all of the way down, we find that $p_1 \cdots p_k$ of the time, the multi-stage trial ends at the leaf we were interested in. So, in this case, we find that each outcome of our 3 coin flip occurs 1/8 =(1/2)(1/2)(1/2) of the time.

For example (c), assuming both the die and the wheel are fair, each outcome for each stage has a 1/4 chance of occuring. So the outcomes on the 2nd row occur 1/16 of the time, and each outcome on the 3rd row occurs 1/64 of the time.

This allows us to compute the probabilities of each individual outcome. To compute the probability of an event, we, as before, add up the probabilities of the outcomes in the event. These two properties are summed up in the

Multiplication Rule: The probability of a leaf is the product of the probabilities along the branches from root to leaf.
Addition Rule: The probability of an event is the sum of the probabilities of the leaves in the event.

For example, in example (c), the probability that we end with a sum of 5 is

$$P(\{123, 132, 141, 232, 241, 32, 341, 41\}) = \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{16} + \frac{1}{64} + \frac{1}{16} = 6\frac{1}{64} + 2\frac{1}{16} = \frac{14}{64} = \frac{7}{32}$$

[Without rolling the die again, just rolling, spinning, and adding, the probability of getting a sum of 5 would be 4(1/16)=1/4. So rerolling makes getting a 5 less likely!]

Conditional Probability:

One of the most interesting aspects of probability is that the probability of an event depends on what you underbarknow. For an ordinary deck of cards, the probability of drawing a 3 from the deck is 1/13 (4 3's out of 52 cards). But if happen to know that the card <u>can't</u> be the 3 of diamonds (because somebody told you it wasn't?) then you know that the probability is really 3/51 (3 3's out of 51 remaining cards), instead. The probability is "conditioned" (changed) by your knowledge.

More generally, $P(A)$ is the relative likelihood that the outcome of a trial is in A. But if we already know that the outcome lies in B (extra knowledge), then to land in A it must <u>really</u> land in $A \cap B$. So given that we know it lies in B, the probability that is in A, which we denote $P(A|B)$ = the probability of A <u>given</u> B, is really the relative size of $A \cap B$ in B, that is, $P(A \cap B)/P(B)$. This is the likehood that, of the outcomes that land in B, we also land in A. Put differently, we can write $P(A \cap B) = P(A|B)P(B)$; if $1/3 = P(B)$ of the time we land in B and, of those times (i.e., given we are in B), $2/5 = P(A|B)$ of the time we land in A, then $P(A \cap B) = (1/3)(2/5) = P(B)P(A|B)$ of the time we land in both A and B.

As an example (work out the tree diagram!), suppose we have two jars, one with 2 white (W) and 3 black (B) marbles, and the other with 4 white and 2 black marbles, and we flip a coin, heads (H) drawing from the first jar and tails (T) from the second. The outcomes are $\{HW, HB, TW, TB\}$, and a tree diagram works out the probabilities. Then

$$P(\text{heads}|\text{black}) = P(\{HB\})/P(\{HB, TB\}) = \frac{(1/2)(3/5)}{(1/2)(3/5) + (1/2)(2/6)} = \frac{9}{14}$$

while $P(\text{heads}) = P(\{HW, HB\}) = (1/2)(2/5) + (1/2)(3/5) = 1/2$. So knowing that we drew a black marble tells us that it is more likely that coin came up heads than we would have otherwise expected. The extra information changed our understanding of the possible outcomes of the coin flip.

The idea of conditional probabilities arise in many situations. One interesting example is medical tests for the presence of disease: a test has a certain probability of saying that you have the disease ("positive") if you have the disease, and a (different) probability of saying that you don't ("negative") if you don't. Ideally, we would like both of these numbers to be high, but usually it is a trade-off between the two, in the design of a test. Suppose that for a given disease, which is known to affect .03 of the population, a test will be positive .99 of the time if you have the disease, and will be negative .98 of the time if you don't. What is the probability that you have the disease, given that you test positive? (The point: whether you have the diesease is "unknown"; we know the test result!). We can model this as a multi-stage "trial": first there is whether or not you have the disease, and second there is the test. We know the probabilities to assign to each branch (build the weighted tree!), and the possible outcomes (DP, DN, dP, dN) [$d =$ no disease], and we compute $P(D|P) = P(DP)/(P(\{DP, dP\}) = ((.03)(.99)/((.03)(.99) + (.97)(.02)) = .604$. So a positive test result only gives a 60% chance of have the disease. So 40% of the time you test positive but do not have the disease ("false positive"). In practice, tests are usually designs to minimze false negatives; you have the disease but the test says no. In this instance, the false negative probability $(D|N)$ is, well, you compute it!

Two events A, B are *independent* if knowledge of one doesn't effect the probability of the other, i.e., $P(A|B) = P(A)$. From our formula, this works out to $P(A \cap B) = P(A)P(B)$. For example, the numbers that come up on two dice when we roll them are independent; knowing one doesn't let us predict the other.

Expected Value:

The whole idea of probability is to predict the likelihood of an event over a large number of trials. If the outcomes are numbers (think: either I pay you something or you pay me!), then we can find the average of the outcomes over a large number of trials: this is the expected value of the trial (or of the probability model).

If a probability model has (numerical) outcomes $x_1, \ldots, x_k$, with probabilities $p_1, \ldots, p_k$, then the expected value of the model, which typically has the unfortunate name of E, is

$$E = p_1x_1 + \cdots p_kx_k$$

The basic idea is that for a large number N of trials, each x_i will occur approximately Np_i times; if we average these numbers, we will be adding N numbers, Np_i of which are x_i, and dividing by N, giving

$$((Np_1)x_1 + \cdots (Np_k)x_k))/N = p_1x_1 + \cdots p_kx_k \text{ for an average.}$$

As a motivational example for what comes next, think of a poll, where the outcomes are yes,no. Think of them as being 1 and 0, instead. Then if 80%, say, of the population would answer yes, then if we make asking an individual a trial, the probability of "yes" is .8, and the expected value of the trial is .8 . This doesn't mean that for a large number of trials we will see an average outcome of .8, just (as the name implies) that we expect to! Our main question, now, is to decide "how often will a large number of trials result in an average outcome close to (far from?) the expected value?" Or put slightly differently, how much can we trust that if we were to run a large number of trials, the average value of out outcomes would turn out to be close to the theoretical expected value?